IN THE SPECIFICATION

[1004] In a hands-free environment, an external speaker may be used. The external speaker and the microphone are usually set far apart, resulting in a longer delay of the echo. In such a case, the adaptive filter 124 may require 512 taps to keep track of the acoustic echo channel 120. The adaptive filter 124 may be used to learn the acoustic echo channel 120 to produce an echo error signal e1(n). The error signal e1(n) is in general a delayed version of far end speech f(t). The input audio picked up by the microphone 128 is passed through an analog to digital converter (ADC) 127. The ADC process may be performed with a limited bandwidth, for example 8 kHz. The digital input signal S(n) is produced. A summer 126 subtracts the echo error signal e1(n) from the input signal S(n) to produce the echo free input signal d(n). When the adaptive filter 124 operates to produce a matched acoustic echo channel, the estimated echo error signal e1(n) is equal to the real echo produced in the acoustic echo channel 120, thus:

$$d(n) = s(n) - e1(n) = [n(n) + e(n)] - e1(n) = n(n)$$

where n(n) and e(n) are discrete-time version of n(t) and e(t) respectively after 8KHz ADC. A voice decoder 123 may produce the far end speech signal f(n) and passed on to an ~~ADC~~DAC 122 to produce the signal f(t). Moreover, the signal d(n) is also passed on to a voice encoder 125 for transmission to the far end user.

[1027] In addition, in accordance with an embodiment, a network VR server ~~206~~ 203 may in communication with base station 202 directly may receive and transmit data exclusively related to VR processing. Server ~~206~~ 203 may perform the back-end VR processing as requested by remote station 201. Server ~~206~~ 203 may be a dedicated server to perform back-end VR processing. An application program user interface (API) provides an easy mechanism to enable applications for VR running on the remote device. Allowing back-end processing at the sever ~~206~~ 203 as controlled by remote device 201 extends the capabilities of the VR API for being accurate,

and performing complex grammars, larger vocabularies, and wide dialog functions. This may be accomplished by utilizing the technology and resources on the network as described in various embodiments.

[1028] A correction to a result of back end VR processing performed at VR server ~~206~~ 208 may be performed by the remote device, and communicated quickly to advance the application of the content data. If the network, in the case of the cited example, returns "Bombay" as the selected city, the user may make correction by repeating the word "Boston." The word "Bombay" may be in an audio response by the device. The user may speak the word "Boston" before the audio response by the device is completed. The input voice data in such a situation includes the names of two cities, which may be very confusing for the back end processing. However, the back end processing in this correction response may take place on the remote device without the help of the network. In alternative, the back end processing may be performed entirely on the remote device without the network involvement. For example, some commands (such as spoken command "STOP" or keypad entry "END") may have their back end processing performed on the remote device. In this case, there is no need to use the network for the back end VR processing, therefore, the remote device performs the front end and back end VR processings. As a result, the front end and back end VR processings at various times during a session may be performed at a common location or distributed.

[1030] Referring to FIG. 5, various blocks of an enhanced echo cancellation system 400 is shown in accordance with various embodiments of the invention. A speaker 401 outputs the audio response of an audio signal 411. The bandwidth of the audio signal 411 is limited in accordance with various aspects of the invention. For example, the bandwidth may be limited to zero to 4 kHz. Such a bandwidth is sufficient for producing a quality audio response from the speaker 401 for human ears. The audio signal 411 may be generated from different sources. For example, the audio signal 411 may be originated from a far end user in communication with a near end user of the device or a voice prompt in an interactive VR system utilized by the device. The far end audio signal f(n) 495 in digital domain may be processed in an ~~ADC 499~~ DAC 499 with a limited bandwidth in accordance with various aspects of the invention. The far end signal 411 with a limited bandwidth is produced. For example, if the sampling frequency of the ~~ADC 499~~ DAC 499 is set to 8 kHz, the audio signal 411 may have a bandwidth of approximately 4 kHz. The signal f(n) 495 may have been received from a voice decoder 498. A unit 410 may produce the input to voice decoder 498 in a form of encoded and modulated signal. The unit 410 may include a

controller, a processor, a transmitter and a receiver. The signal decoded by voice decoder 498 may be in a form of audio PCM samples. Normally, the PCM samples data rate is 8K samples per second in traditional digital communication systems. The audio PCM samples are converted to analog audio signal 411 via 8KHz ~~ADC 499~~ DAC 499and the played by speaker 401. The produced audio, therefore, is band limited in accordance with various aspects of the invention.